Docket No.: NEB-228

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

INVENTOR(S): Richard D. Morgan
Paul Walsh

TITLE: A Novel Type II Restriction Endonuclease, CstMI, Obtainable From Corynebacterium striatum M82B And A Process For Producing The Same

ATTORNEY: Gregory D. Williams
General Counsel
(Reg. No. 30901)

NEW ENGLAND BIOLABS, INC.
32 Tozer Road
Beverly, MA 01915
(978) 927-5054 x: 292/373

EXPRESS MAILING LABEL NO.:

Customer No.: 28986

A NOVEL TYPE II RESTRICTION ENDONUCLEASE, CstMI, OBTAINABLE FROM *Corynebacterium striatum* M82B AND A PROCESS FOR PRODUCING THE SAME

BACKGROUND OF THE INVENTION

The present invention relates to a novel type II restriction endonuclease, CstMI. CstMI consists of one polypeptide which possesses two related enzymatic functions. CstMI is an endonuclease that recognizes the DNA sequence 5'-AAGGAG-3' and cleaves the phosphodiester bond between the 20th and 21st residues 3' to this recognition sequence on this DNA strand, and between the 18th and 19th residues 5' to the recognition sequence on the complement strand 5'-CTCCTT-3' to produce a 2 base 3' extension (hereinafter referred to as the CstMI restriction endonuclease). CstMI has a second enzymatic activity that recognizes the same DNA sequence, 5'-AAGGAG-3', but modifies this sequence by the addition of a methyl group to prevent cleavage by the CstMI endonuclease. The present invention also relates to the DNA fragment encoding the CstMI enzyme, a vector containing this DNA fragment, a transformed host containing this DNA fragment, and a process for producing CstMI restriction endonuclease from such a transformed host. CstMI was identified as a potential endonuclease because of its amino acid sequence similarity to MmeI (see U.S. Application Serial No. _______________, filed concurrently herewith).

Restriction endonucleases are a class of enzymes that occur naturally in prokaryotes. There are several classes of restriction systems known, of which the type II endonucleases are the class useful in genetic engineering. When these type II endonucleases are purified away from other contaminating prokarial

components, they can be used in the laboratory to break DNA molecules into precise fragments. This property enables DNA molecules to be uniquely identified and to be fractionated into their constituent genes. Restriction endonucleases have proved to be indispensable tools in modern genetic research. They are the biochemical 'scissors' by means of which genetic engineering and analysis is performed.

Restriction endonucleases act by recognizing and binding to particular sequences of nucleotides (the 'recognition sequence') along the DNA molecule. Once bound, the type II endonucleases cleave the molecule within, or to one side of, the sequence. Different restriction endonucleases have affinity for different recognition sequences. The majority of restriction endonucleases recognize sequences of 4 to 6 nucleotides in length, although recently a small number of restriction endonucleases which recognize 7 or 8 uniquely specified nucleotides have been isolated. Most recognition sequences contain a dyad axis of symmetry and in most cases all the nucleotides are uniquely specified. However, some restriction endonucleases have degenerate or relaxed specificities in that they recognize multiple bases at one or more positions in their recognition sequence, and some restriction endonucleases recognize asymmetric sequences. *Hae*III, which recognizes the sequence 5'-GGCC-3', is an example of a restriction endonuclease having a symmetrical, non-degenerate recognition sequence; *Hae*II, which recognizes 5'-(Pu)GCGC(Py)-3' typifies restriction endonucleases having a degenerate or relaxed recognition sequence; while *Bsp*MI, which recognizes 5' - ACCTGC - 3' typifies restriction endonucleases having an asymmetric recognition sequence. Type II endonucleases with

symmetrical recognition sequences generally cleave symmetrically within or adjacent to the recognition site, while those that recognize asymmetric sequences tend to cleave at a distance of from 1 to 20 nucleotides to one side of the recognition site. The enzyme of this application, CstMI, (along with MmeI) has the distinction of cleaving the DNA at the farthest distance from the recognition sequence of any known type II restriction endonuclease. More than two hundred unique restriction endonucleases have been identified among several thousands of bacterial species that have been examined to date.

A second component of restriction systems are the modification methylases. These enzymes are complementary to restriction endonucleases and they provide the means by which bacteria are able to protect their own DNA and distinguish it from foreign, infecting DNA. Modification methylases recognize and bind to the same nucleotide recognition sequence as the corresponding restriction endonuclease, but instead of breaking the DNA, they chemically modify one or other of the nucleotides within the sequence by the addition of a methyl group. Following methylation, the recognition sequence is no longer cleaved by the restriction endonuclease. The DNA of a bacterial cell is modified by virtue of the activity of its modification methylase and it is therefore insensitive to the presence of the endogenous restriction endonuclease. It is only unmodified, and therefore identifiably foreign, DNA that is sensitive to restriction endonuclease recognition and cleavage. Modification methyltransferases are usually separate enzymes from their cognate endonuclease partners. In some cases, there is a single polypeptide that possesses both a modification methyltransferase

function and an endonuclease function, for example, Eco57I. In such cases, there is usually a second methyltransferase present as part of the restriction-modification system. CstMI, however, consists of a single polypeptide that possesses both a modification methyltransferase function and an endonuclease function but does not have a second methyltransferase peptide as part of the restriction modification system. In this regard CstMI is similar to the MmeI restriction modification system.

Endonucleases are named according to the bacteria from which they are derived. Thus, the species *Haemophilus aegyptius*, for example synthesizes 3 different restriction endonucleases, named *Hae*I, *Hae*II and *Hae*III. These enzymes recognize and cleave the sequences 5'-(W)GGCC(W)-3', 5'-(Pu)GCGC(Py)-3' and 5'-GGCC-3' respectively. *Escherichia coli* RY13, on the other hand, synthesizes only one enzyme, *Eco*RI, which recognizes the sequence 5'-GAATTC-3'.

While not wishing to be bound by theory, it is thought that in nature, restriction endonucleases play a protective role in the welfare of the bacterial cell. They enable bacteria to resist infection by foreign DNA molecules such as viruses and plasmids that would otherwise destroy or parasitize them. They impart resistance by binding to infecting DNA molecules and cleaving them in each place that the recognition sequence occurs. The disintegration that results inactivates many of the infecting genes and renders the DNA susceptible to further degradation by exonucleases.

More than 3000 restriction endonucleases have been isolated from various bacterial strains. Of these, more than 240 recognize unique sequences, while the rest

share common recognition specificities. Restriction endonucleases which recognize the same nucleotide sequence are termed "isoschizomers." Although the recognition sequences of isoschizomers are the same, they may vary with respect to site of cleavage (e.g., *Xma*I v. *Sma*I, Endow, et al., *J. Mol. Biol.* **112**:521 (1977); Waalwijk, et al., *Nucleic Acids Res.* **5**:3231 (1978)) and in cleavage rate at various sites (*Xho*I v. *Pae*R7I, Gingeras, et al., *Proc. Natl. Acad. Sci.* U.S.A. **80**:402 (1983)).

Restriction endonucleases have traditionally been classified into three major classes; type I, type II and type III. The type I restriction systems assemble a multi-peptide complex consisting of restriction polypeptide, modification polypeptide, and specificity, or DNA recognition, polypeptide. Type I systems require á divalent cation, ATP and S-adenylosyl-methionine (SAM) as cofactors. Type I systems cleave DNA at random locations up to several thousand basepairs away from their specific recognition site. The type III systems generally recognize an asymmetric DNA sequence and cleave at a specific position 20 to 30 basepairs to one side of the recognition sequence. Such systems require the cofactor ATP in addition to SAM and a divalent cation. The type III systems assemble a complex of endonuclease polypeptide and modification polypeptide that either modifies the DNA at the recognition sequence or cleaves. Type III systems produce partial digestion of the DNA substrate due to this competition between their modification and cleavage activities, and so have not been useful for genetic manipulation.

CstMI can be classified as a type II endonuclease in that it does not require ATP for DNA cleavage

activity. Unlike other type II enzymes, however, CstMI consists of a single polypeptide that combines both endonuclease and modification activities and is sufficient by itself to form the entire restriction modification system. CstMI, like the related endonuclease MmeI, cleaves the farthest distance from the specific DNA recognition sequence of any type II endonuclease. CstMI is quite large and appears to have three functional domains combined in one polypeptide. These consist of an amino-terminal DNA cleavage domain which may also be involved in DNA recognition, a DNA modification domain most similar to the gamma-class N6mA methyltransferases, and a carboxy-terminal domain presumed to be involved in dimer formation and possibly DNA recognition. The enzyme requires SAM for both cleavage and modification activity. The single CstMI polypeptide is sufficient to modify the plasmid vector carrying the gene *in vivo* to provide protection against CstMI cleavage *in vitro*, yet it is also able to cleave unmodified DNAs *in vitro* when using the endonuclease buffer containing Mg++ and SAM.

There is a continuing need for novel type II restriction endonucleases. Although type II restriction endonucleases which recognize a number of specific nucleotide sequences are currently available, new restriction endonucleases which recognize novel sequences provide greater opportunities and ability for genetic manipulation. Each new unique endonuclease enables scientists to precisely cleave DNA at new positions within the DNA molecule, with all the opportunities this offers.

SUMMARY OF THE INVENTION

In accordance with the present invention, there is provided a novel DNA fragment encoding a novel restriction endonuclease, obtainable from *Corynebacterium striatum* M82B (GenBank Accession #AAG03371) or from the transformed E. coli strain NEB#1530. The endonuclease is hereinafter referred to as "CstMI", which endonuclease:

(1) recognizes the nucleotide sequence 5'-AAGGAG-3' in a double-stranded DNA molecule as shown below,

5'-AAGGAG-3'
3'-TTCCTC-5'

(wherein G represents guanine, C represents cytosine, A represents adenine and T represents thymine;

(2) cleaves DNA in the phosphodiester bond following the 20th nucleotide 3' to the recognition sequence 5'-AAGGAG-3 and preceding the 18th nucleotide 5' to the recognition sequence in the complement strand of 5'-CTCCTT-3' to produce a 2 base 3' extension:
5'-AAGGAG(N20)/-3'
3'-TTCCTC(N18)/-5'; and

(3) methylates the recognition sequence specified in (1) *in vivo* to protect the host DNA from cleavage by the CstMI endonuclease activity;

The present invention further relates to a process for the production of the restriction endonuclease CstMI. This process comprises culturing a transformed host, such as *E. coli*, containing the DNA fragment encoding the CstMI restriction system polypeptide, collecting the cultured cells, obtaining a cell-free extract therefrom and separating and collecting the restriction endonuclease CstMI from the cell-free extract.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 - Agarose gel showing CstMI cleavage of lambda, T7, phiX174, pBR322 and pUC19 DNAs. Lanes 1, 7, 13 and 20: lambda-HindIII, PhiX174-HaeIII size standards; lane 2: pUC19 DNA + CstMI + Eco0109I; lane 3: pUC19 DNA + CstMI + PstI; lane 4: pUC19 DNA + CstMI + AlwNI; lane 5: pUC19 DNA + CstMI + XmnI; lane 6: pUC19 DNA + CstMI; lane 8: pBR322 DNA = CstMI + ClaI; lane 9: pBR322 DNA + CstMI + NruI; lane 10: pBR322 DNA + CstMI + NdeI; lane 11, pBR322 DNA + CstMI + PstI; lane 12: pBR322 DNA + CstMI; lane 14: PhiX174 DNA = CstMI + PstI; lane 15: PhiX174 DNA + CstMI + SspI; lane 16: PhiX174 DNA + CstMI + NciI; lane 17: PhiX174 DNA + CstMI + StuI; and lane 18: PhiX174 DNA + CstMI

Figure 2 - DNA sequence of the CstMI gene locus (SEQ ID NO:1).

Figure 3 - Amino acid sequence of the CstMI gene locus (SEQ ID NO:2).

Figure 4 - Agarose gel showing CstMI protection of pTBCstMI.3 DNA and cleavage of unmodified DNA substrate. lane 1 and 5: lambda-HindIII, PhiX174-HaeIII size

standards; lane 2: pTBCstMI.3 + Eco0109I; lane 3: pTBCstMI.3 + eco0109I + CstMI; lane 4: pTBCstMI.3 + CstMI + pUC19 DNA.

Figure 5 - Determination of the CstMI cleavage site.

Figure 5A: Location of cleavage on 5'AAGGAG-3" strand (SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:5).

Figure 5B: location of cleavage on 5'-CTCCTT-3' strand (SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8).

Figure 6 - Sequence alignment of CstMI (SEQ ID NO:9) and MmeI (SEQ ID NO:10) amino acid sequences

Figure 7 - Photograph depicting titer of CstMI crude extract on lambda DNA. The reaction mixture is NEBuffer 4 supplmented with 100 uM SAM with 1 Ug lambda DNA per 50 uL. Digestion took place at 37°C for one hour. Lane 1 - lambda-HindIII and PhiX174-HaeIII marker; Lane 2 - 8 uL crude extract/50 uL reaction mix; Lane 3 - 4 uL crude extract/50 uL reaction mix; Lane 4 - 2 uL crude extract/50 uL reaction mix; Lane 5 - 1 uL crude extract/50 uL reaction mix; Lane 6 - 0.5 uL crude extract/50 uL reaction mix; Lane 7 - 0.25 uL crude extract/50 uL reaction mix.

DETAILED DESCRIPTION OF THE INVENTION

The MmeI endonuclease was cloned New England Biolabs, Inc. (Beverly, MA) and its amino acid sequence was determined (U.S. Application Serial No. _____, filed concurrently herewith, the disclosure of which is herein incorporated by reference). A BLAST search of the Genbank database using the MmeI endonuclease amino acid sequence as the query returned a number of sequences

that were highly significantly similar to MmeI. Among these was a sequence, GenBank accession #AAG03371, which encoded a gene labeled gcrY, and annotated as a "hypothetical 107.5 kDa protein". This hypothetical protein was encoded on a 51,409 base pair plasmid isolated from *Corynebacterium striatum* M82B (see Tauch,A., Krieft,S., Kalinowski,J. and Puhler,A., "The 51,409-bp R-plasmid pTP10 from the multiresistant clinical isolate Corynebacterium striatum M82B is composed of DNA segments initially identified in soil bacteria and in plant, animal, and human pathogens" Mol. Gen. Genet. 263 (1), 1-11 (2000)). A sample of this plasmid DNA was kindly provided by the author, Andreas Tauch. The DNA sequence encoding and flanking the potential endonuclease gene was known. Primers were designed to specifically amplify the gene from *Corynebacterium striatum* M82B DNA, with convenient restriction enzyme sites added to facilitate cloning into a vector. The amplified gene was inserted into an expression vector and cloned into *an E. coli* host. Transformed host cells were tested and several were found to express an endonuclease activity when incubated in NEBuffer 4 supplemented with 100μM SAM (S-adenosyl-methionine) (Figure 7). The DNA recognition sequence of this new endonuclease was determined by mapping the positions of cleavage in pUC19, pBR322 and PhiX174 DNAs. These locations of cleavage were found to be consistent with the sequence 5'-AAGGAG-3' (or 5'-CTCCTT-3' on the complement DNA strand). This novel enzyme was named CstMI (from **C**orynebacterium **st**riatum **M**82B). This recognition sequence is quite different from that of MmeI, which recognizes 5'-TCC(Pu)AC-3', even though the enzymes share approximately 40% identical and 51% similar amino acids in their sequences (Figure 8). The point of DNA cleavage relative to the recognition

sequence was determined by cutting an appropriate DNA with CstMI, purifying the DNA and subjecting it to standard dideoxy automated sequencing. CstMI was found to cleave DNA at the same position relative to its recognition sequence as MmeI; namely after the 20th nucleotide 3' to the 5'-AAGGAG-3' recognition sequence strand, and before the 18th nucleotide 5' to the 5'-CTCCTT-3' recognition sequence strand, producing a 2 base pair 3' extension. CstMI was also found to *in vivo* modify the recombinant expression vector, pTBCstMI.3, such that it was protected against CstMI endonuclease activity *in vitro*.

In Example I below we describe the cloning and expression of CstMI.

In Example II we obtained CstMI by culturing a transformed host carrying the CstMI gene, such as *E. coli* ER2683 carrying pTBCstMI.3 and recovering the endonuclease from the cells. A sample of *E. coli* ER2683 carrying pTBCstMI.3 (NEB#1530) has been deposited under the terms and conditions of the Budapest Treaty with the American Type Culture Collection (ATCC) on ____________, 2003 and bears the ATCC Accession No.____.

For recovering the enzyme of the present invention *E. coli* carrying pTBCstMI.3 (NEB#1530) may be grown using any suitable technique. For example, *E. coli* carrying pTBCstMI.3 may be grown in Luria broth media containing 100 μg/ml ampicillin and incubated aerobically at 37°C with aeration. Cells in the late logarithmic stage of growth are induced by adding 0.3mM IPTG, grown for an additional 4 hours, collected by centrifugation and either disrupted immediately or stored frozen at -70°C.

The CstMI enzyme can be isolated from *E. coli* carrying pTBCstMI.3 cells by conventional protein purification techniques. For example, cell paste is suspended in a buffer solution and treated by sonication, high pressure dispersion or enzymatic digestion to allow extraction of the endonuclease by the buffer solution. Intact cells and cellular debris are then removed by centrifugation to produce a cell-free extract containing CstMI. The CstMI endonuclease, along with its corresponding intrinsic methylase activity, is then purified from the cell-free extract by ion-exchange chromatography, affinity chromatography, molecular sieve chromatography, or a combination of these methods to produce the endonuclease of the present invention.

The present invention is further illustrated by the following Examples. These Examples are provided to aid in the understanding of the invention and are not construed as a limitation thereof.

The references cited above and below are herein incorporated by reference.

EXAMPLE I

CLONING THE CstMI ENDONUCLEASE

1. Identifying the CstMI endonuclease gene from *Corynebacterium straitum* M82B 51,409 bp plasmid pTP10 DNA: The putative CstMI endonuclease open reading frame was identified by a BLAST search of the nonredundant sequences in the GenBank database. The BLAST algorithm was performed using the MmeI amino acid sequence as the query, with parameters of word size = 3, matrix =

BLOSUM62, gap costs of 11 for the existence of a gap and 1 for an extension of a gap, with no masking for low complexity. The open reading for the CstMI endonuclease, found in the *Corynebacterium straitum* M82B 51,409 bp plasmid pTP10 DNA, labeled gcrY and annotated as a 'hypothetical protein," yielded a very highly significant expectation value of $E=e^{-171}$, making it an excellent candidate for a new MmeI-like endonuclease.

2. DNA purification: A DNA preparation of the *Corynebacterium straitum* M82B 51,409 bp plasmid pTP10 was kindly supplied by Andreas Tauch.

3. Cloning the CstMI open reading frame: Oligonucleotide primers were synthesized to specifically amplify the CstMI gene from *Corynebacterium striatum* pTP10 plasmid DNA for expression in the cloning vector pRRS (Skoglund, Gene 88:1-5 (1990)). The forward primer contained a NsiI site for cloning, a stop codon in frame with the lacZ gene of the vector, a consensus E. coli ribosome binding site, the ATG start codon for translation and 20 nucleotides that matched the *Corynebacterium striatum* pTP10 plasmid DNA sequence at the beginning of the CstMI open reading frame:

CstMI expression primer forward (#282-48):
5' - GTTATGCATTTAAGGAGGTAACATATGGTTATGGCCCCTACGAC-3'
(SEQ ID NO:11)

The reverse primer contained a BamHI for cloning and 21 nucleotides that matched the the *Corynebacterium striatum* DNA sequence beginning at the C base in the complement strand corresponding to the G base of the stop codon TAG of the CstMI open reading frame:

CstMI expression primer reverse (#282-49):
5' -GTTGGATCCTCGAGGGCAAGACATATCAAGCCTTC -3'
(SEQ ID NO:12)

The CstMI gene was amplified in a PCR reaction by combining:

50 µl 10X Thermopol buffer (NEB)
30 µl 4mM dNTP solution
12.5 µl forward primer #282-48 (10µM stock)
12.5 µl reverse primer #282-49 (10µM stock)
5 µl *Corynebacterium striatum* pTP10 plasmid DNA (5µg/ml stock)
387 µl dH_2O
3 µl (6 units) Vent® DNA polymerase

The reaction was mixed and aliquoted into 5 tubes of 80 µl each. $MgSO_4$ was added (100mM stock) to bring the final concentration of Mg++ ions to 2mM, 3mM, 4mM, 5mM and 6mM respectively. The cycling parameters were 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 3 minutes, for 5 cycles, followed by 23 cycles of 95°C for 30 seconds, 64°C for 30 seconds, 72°C for 3 minutes. The reactions were analyzed by gel electrophoresis and the 2mM through 5mM Mg++ reactions were found to contain a DNA band of the desired size of 2.9kb. These reactions were pooled and the 2.9kb band was gel purified. The 2.9kb amplified CstMI gene fragment was digested with BamHI and NsiI endonucleases (NEB) in the following reaction conditions:

2 µl 10X BamHI reaction buffer (NEB)
8 µl CstMI gene 2.9 kb amplified DNA fragment
10 µl dH2O
0.5 µl BamHI endonuclease (10 units)

0.5 μl PstI endonuclease (10 units)

The reaction was mixed and incubated for 1 hour at 37°C. The endonucleases were heat killed by incubating at 80°C for 20 minutes.

The cleaved CstMI gene DNA fragment was ligated to the pRRS vector. 10 μl of the digested, purified 2.9kb CstMI fragment was combined with 3 μl pRRS vector previously cleaved with BamHI and PstI and purified, 5 μl dH_2O, 2 μl 10X T4 DNA Ligase Buffer (NEB), the reaction was mixed, and 1 μl of T4 DNA Ligase was added. The reaction was incubated at 16°C for 16 hours. 5 μl of the ligation reaction was transformed into 100 μl electro-competent *E. coli* ER2683 cells, the cells were grown out in 1 ml Luria broth for 45 minutes, then 20 μl and 200 μl were plated on L-broth plates containing 100 $\mu g/ml$ ampicillin and incubated at 37°C overnight. Approximately 100 transformants were obtained and 4 representatives were analyzed as follows: plasmid from each colony was isolated by miniprep procedures and digested with PvuII endonucleases to determine if they contained the correct size insert. 3 of the 4 transformants had the correct size insert of approximately 2900 bp. The 3 insert containing clones were digested with MmeI endonuclease to see if this open reading frame produced an enzyme that recognized the same sequence as MmeI and thus protected the plasmid DNA of the clone from MmeI digestion. All three clones were cut with MmeI endonuclease, indicating that this enzyme did not modify the DNA at the MmeI recognition site.

Two of the clones were tested to see if they produced any endonuclease activity. The purified plasmid DNAs were transformed into E. coli strain ER2796. 6

colonies that grew up from one of the clones were tested for endonuclease activity. The six colonies were inoculated into 50 ml luria broth containing 100 µg/ml ampicillin and grown overnight at 37°C with shaking. The cells were then harvested by centrifugation, resuspended in 1.5 ml buffer (20 mM Tris-HCl, 1 mM DTT, 0.1 mM EDTA) and lysed by sonication. The lysate was assayed for endonuclease activity by serial dilution of the lysate in 1X reaction buffer NEBuffer 4 (New England Biolabs) containing 20 µg/ml lambda DNA substrate and supplemented with SAM at 80 µM final concentration. The reactions were incubated for 1 hour at 37°C. The reaction products were analyzed by agarose gel electrophoresis on a 1% agarose gel in 1X TBE buffer. Two of the six clones clearly had endonuclease activity, three did not show endonuclease activity and one appeared to be a contaminant (not an E. coli clone). The most active clone was designated strain NEB#1530 and was used for subsequent production of CstMI. The plasmid construct expressing CstMI activity in this clone was designated pTBCstMI.3.

EXAMPLE II

PRODUCTION OF CstMI ENDONUCLEASE

A single colony of *E. coli* ER2683 carrying the CstMI gene in the vector pTBCstMI.3 (NEB#1530) was grown in 2 liter of Luria broth. The cells were grown aerobically at 37°C for 14 hours, then IPTG was added to 0.3mM final concentration and the cells were grown for 2 more hours. The cells were collected by centrifugation, yielding two grams of wet cell pellet.

4 grams of CstMI expressing NEB#1530 cell pellet was suspended in 10 milliliters of Buffer A (20 mM Tris-

HCl (pH 8.0), 50 mM NaCl, 1.0 mM DTT, 0.1 mM EDTA) and sonicated for 6 minutes at a 50% pulse to disrupt the cells. The lysate was centrifuged at ~30,000 x G for 15 minutes and the supernatant collected. (Figure 7)

The supernatant solution was applied to a 8 ml Heparin Hyper-D column (BioSepra SA) which had been equilibrated in buffer A. A 16 mL wash of buffer A was applied, then a 150 mL gradient from 0.05M to 1M NaCl in buffer A was applied and 3 mL fractions were collected. Fractions were assayed for CstMI endonuclease activity by incubating with 1 μg Lambda DNA (NEB) in 50 μl NEBuffer 4, supplemented with 100 μM S-adenosyl-L-methionine (SAM) for 15 minutes at 37° C. CstMI activity eluted at 0.33M to 0.44M NaCl.

The Heparin Hyper-D column fractions containing the CstMI activity were pooled, diluted to 50mM NaCl with buffer A (without NaCl) and applied to a 3 ml Heparin-TSK column (TosoHaas) which had been equilibrated with buffer A. A wash of 6 ml buffer A was applied, followed by a 50 ml linear gradient of NaCl from 0.05M to 1.0M in buffer A. Fractions were collected and assayed from CstMI endonuclease activity. The CstMI activity eluted between 0.44 M and 0.48 M NaCl.

The Heparin-TSK column fractions containing CstMI activity were pooled, diluted to 50mM NaCl with buffer A (without NaCl) and applied to a 1 ml Mono-Q FPLC column (Pharmacia) equilibrated with buffer A. A wash of 2 ml buffer A was applied, followed by a 40 ml linear gradient of NaCl from 0.05 M to 0.6 M in buffer A. 1 ml fractions were collected and assayed from CstMI endonuclease activity. CstMI eluted from 0.28 M to 0.4 M NaCl. The purified CstMI fractions were pooled (4 ml)

and dialyzed against storage buffer (10 mM Tris (pH 7.9), 50 mM KCl, 1mM DTT, 0.1 mM EDTA, 50% glycerol). The purified CstMI enzyme was stored at -20°C. The CstMI endonuclease obtained was substantially pure.

Activity determination:

Samples from 1-4 µl were added to 50 µl substrate solution consisting of 1X NEBuffer 4, 100 µM S-adenosyl-L-methionine, and 1 µg DNA (lambda, PhiX174, pBR322 or pUC19 DNAs). Reactions were incubated for 15 minutes at 37°, received 20 µl stop solution and were analyzed by electrophoresis on a 1% agarose gel (figure 1)

EXAMPLE III

DETERMINATION OF THE CstMI ENDONUCLEASE CLEAVAGE SITE

The location of CstMI cleavage relative to the recognition sequence was determined by cleaving a suitable DNA molecule and then performing DNA sequencing from a suitable primer to the end of the cleaved DNA template. In this example pUC19 DNA and pBR322 DNA were employed as the template. These templates were chosen because there were CstMI sites in both orientations located within several hundred base pairs from standard sequencing primers. Any sequenceable DNA that has a CstMI site within several hundred base pairs of a priming site will work for this analysis, however. The pUC19 DNA was cleaved with CstMI by combining:

50 µl 10X NEBuffer #4
15 µl pUC19 DNA (15 µg)
435 µl dH_2O
20 µl CstMI (fraction 28 off the MonoQ column)

and incubating for 15 minutes at 37°C. pBR322 DNA and pUC19-Adeno2 BC4 DNAs were cut using the same conditions. The cleaved DNAs were purified and concentrated using a Qiagen QiaPrep DNA spin column according to the manufacturer's instructions. The DNAs were eluted in a volume of 100 µl.

Sequencing Reactions

The sequencing reactions were performed using an ABI377 DNA sequencer according to the manufacturer's instructions. The cleaved pUC19 DNA was sequenced with primers NEB1233 and NEB1238 (New England Biolabs) to examine the cut at position 240:

NEB1233 5'-AGCGGATAACAATTTCACACAGGA-3' (SEQ ID NO:13)
NEB1238 5'-CCTATAAAAATAGGCGTATCACGAGGCCCT-3 (SEQ ID NO:14)

The cleaved pBR322 DNA was sequenced with primers NEB1242 and NEB1247 (New England Biolabs) to examine the cut at 537.

NEB1242: 5'-AAGTGCGGCGACGATAGTCATGCCCCGCGC-3' (SEQ ID NO:15)
NEB1247: 5'-TACTTGGAGCCACTATCGACTACGCGATCA-3' (SEQ ID NO:16)

A pUC19-derived plasmid (pUC19-Adeno2 BC4) that contains a fragment of Adeno2 DNA from BstBI (10,670) to ClaI (18,657) inserted at the AccI site of pUC19 was also cut with CstMI and sequenced with primer NEB1224 to examine the CstMI site of Adeno2 DNA at 10,743.

NEB1224: 5'-CGCCAGGGTTTTCCCAGTCACGAC-3' (SEQ ID NO:17)

The results indicate CstMI cleaves DNA between the 20th and the 21th nucleotides 3' to the recognition sequence 5'-AAGGAG-3 in this DNA strand, and between the 18th and 19th nucleotides 5' to the recognition sequence in the complement stand, 5'-CTCCTT-3', to produce a 2 base 3' extension (Figure 5).

EXAMPLE IV

THE CstMI ENDONUCLEASE PROVIDES IN VIVO PROTECTION AGAINST CstMI CLEAVAGE

The plasmid pTBCstMI.3 was purified from NEB#1530 using the Qiagen miniprep protocol. This plasmid has two CstMI sites in the vector backbone, and two site within the CstMI gene. The plasmid was digested with CstMI to test whether this DNA was resistant to CstMI endonuclease activity, which would indicate that the single CstMI gene was able to methylate DNA *in vivo* to protect the host DNA against its endonuclease activity. To test this the following were combined:

6 µl pTBCstMI.3 plasmid DNA
15 µl 10X NEBuffer 4
0.5 µl SAM (32mM stock solution)
129 µl dH2O
3 µl Eco0109I endonuclease (to linearize the plasmid)

The reaction mix was split into two pools, one of 50 µl, to which nothing more was added, and one of 100 µl, to which CstMI endonuclease was added. The CstMI containing reaction was then split into two equal portions and 0.5 µl of pUC19 DNA (0.5 µg) was added to

one half as a positive control for CstMI endonuclease activity (Figure 4).